DOCUMENT RESUME

ED 212 662 TM 820 070

AUTHOR Fuchs, Lynn; Deno, Stanley
TITLE The Relationship between Curriculum-Based Mastery Measures and Standardized Achievement Tests in Reading.
INSTITUTION Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY Office of Special Education (ED), Washington, D.C.
REPORT NO IRLD-RR-57
PUB DATE Aug 81
CONTRACT 300-80-0622
NOTE 51p.
AVAILABLE FROM Editor, IRLD, 350 Elliott Hall, 75 East River Road, University of Minnesota, Minneapolis, MN 55455 ($3.00).

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Basal Reading; Elementary Education; Learning Disabilities; *Oral Reading; *Reading Achievement; Reading Tests; *Standardized Tests; Student Placement; Testing; *Test Validity
IDENTIFIERS *Curriculum Related Testing; Test Curriculum Overlap

ABSTRACT

The edumetric adequacy of curriculum-based reading measures were examined for two basal reading programs. On the basis of reading aloud performance, 91 elementary students were assigned seven instructional placement scores within each basal series. Students also were measured on standardized reading achievement tests. Generally, correlations between instructional scores within each series and performance on standardized tests were high and similar, providing evidence that the curriculum-based reading measures are valid with respect to technically adequate standardized tests; however, validity was dependent on the placement criteria employed. Additional analysis revealed other important edumetric effects of using different placement criteria. The technical adequacy of curriculum-based reading measurement is discussed along with recommendations for developing instructionally useful measurement procedures. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED212662

University of Minnesota

Research Report No. 57

THE RELATIONSHIP BETWEEN CURRICULUM-BASED MASTERY MEASURES AND STANDARDIZED ACHIEVEMENT TESTS IN READING

Lynn Fuchs and Stanley Deno

IRLD

Institute for Research on Learning Disabilities

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 820070

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Ysseldyke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IRLD

Director: James E. Ysseldyke

Associate Director: Phyllis K. Mirkin

The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with the Office of Special Education, Department of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

During 1980-1983, Institute research focuses on four major areas:

- Referral
- Identification/Classification
- Intervention Planning and Progress Evaluation
- Outcome Evaluation

Additional information on the Institute's research objectives and activities may be obtained by writing to the Editor at the Institute (see Publications list for address).

The research reported herein was conducted under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent the official position of the Office of Special Education.

Research Report No. 57

THE RELATIONSHIP BETWEEN CURRICULUM-BASED MASTERY MEASURES AND STANDARDIZED ACHIEVEMENT TESTS IN READING

Lynn Fuchs and Stanley Deno

Institute for Research on Learning Disabilities

University of Minnesota

August, 1981

Abstract

The edumetric adequacy of curriculum-based reading measures was examined for two basal reading programs. On the basis of reading aloud performance, 91 elementary students were assigned seven instructional placement scores within each basal series. Students also were measured on standardized reading achievement tests. Generally, correlations between instructional scores within each series and performance on standardized tests were high and similar, providing evidence that the curriculum-based reading measures are valid with respect to technically adequate standardized tests; however, validity was dependent on the placement criteria employed. Additional analyses revealed other important edumetric effects of using different placement criteria. The technical adequacy of curriculum-based reading measurement is discussed along with recommendations for developing instructionally useful measurement procedures.

The Relationship Between Curriculum-based Mastery Measures and Standardized Achievement Tests in Reading

Since the passage of P.L. 94-142 in 1975, and with the increasing demand for accountability in the schools, educators have been required to support empirically their decisions that affect handicapped children. Because it provides the essential documentation and information for making decisions about pupils, measurement represents a critical component of the educational process (Glaser & Nitko, 1971; Ysseldyke, 1979). Unfortunately, the measurement systems available for gathering assessment information fail to provide all of the essential decision-making data while maintaining demonstrated technical adequacy (cf. Thorndike, 1971). Therefore, the data base on which educational decisions are made is typically less than satisfactory.

Pre-post administration of norm-referenced achievement tests is the most commonly employed measurement format (Anastasi, 1976; Glaser & Nitko, 1971; Tyler, 1951). Yet, norm-referenced achievement tests suffer from the poor reliability characteristic of difference scores (Stanley, 1971), are unsuitable for ongoing monitoring of the appropriateness of educational programs (Jenkins, Deno, & Mirkin, 1979), and frequently lack content validity with respect to a student's curriculum (Armbruster, Stevens, & Rosenshine, 1977; Eaton & Lovitt, 1972; Jenkins & Pany, 1978). Following is a brief discussion of these limitations and of alternative measurement formats that might revolve these problems and improve the data base on which educational decisions are made.

Limitations of Pre-Post Testing on Standardized Achievement Tests

Pre-post testing on standardized achievement tests is of limited use

for the purpose of making educational programming decisions. Stanley (1971) demonstrated that pre and post testing on the same or similar tests leads to low reliability of an examinee's difference score. When the correlation is high between pre- and posttests, there is a great overlap between the true scores of the examinees; a high proportion of the obtained - true score difference is error. Additionally, when pre-post testing is employed, a determination of program effectiveness is made at the end of the treatment period. This summative evaluation prevents the educator from employing the measurement data to improve the student's program throughout the treatment period.

As an alternative, students can be measured at regular, frequent points within the treatment period to formulate ongoing or formative decisions concerning program effectiveness. Unfortunately, achievement tests lack utility for such frequent checks on program effectiveness because (a) they are too long to be administered regularly, and (b) with frequent use, children inadvertently learn tests that typically have a limited number of alternate forms.

Another problem of standardized achievement tests is that they frequently lack content validity with respect to a student's curriculum. Because the content of a basal reader may be unevenly represented in different tests, a student's obtained score may be dependent on the choice of reading test rather than on actual student achievement. Several studies illustrate the potential lack of consonance between a test's content and a pupil's curriculum. For example, Jenkins and Pany (1978) used several popular basal readers and standardized achievement tests to look at the relationship between a student's reading

curriculum and test performance. They computed a hypothetical Grade Equivalent Score (GES) for each test based on the assumption that a student had learned the words taught in a particular reader. The GESs produced by this procedure indicated that achievement tests are differentially sensitive to vocabulary taught in a particular curriculum. For instance, a beginning second grader who had learned the vocabulary in the Macmillan series would score seriously below grade level on the Word Analysis portion of the Metropolitan Achievement Test, but at or above grade level on the other tests. This differential sensitivity of achievement tests undermines their utility in making educational decisions.

Differential sensitivity of achievement tests is not confined to instruments that focus on word recognition skills. Armbruster, Stevens, and Rosenshine (1977) compared the content of two reading achievement tests, the Metropolitan Achievement Test (MAT) and California Achievement Test (CAT), with the content of the reading comprehension exercises in the Economy, Ginn 360, and Houghton-Mifflin reading series. They found that test items of the MAT and CAT failed to cover roughly 64% of Economy, 65% of Ginn, and 79% of Houghton-Mifflin reading comprehension exercises. Furthermore, there were large differences in the relative emphases between the reading series and the achievement tests. Based on the percentage of different types of reading comprehension exercises and the percentage of items tapping these different comprehension skills, the Ginn series correlated .10 with the MAT, while the Houghton-Mifflin series correlated .42 with the CAT.

The findings of Jenkins and Pany and of Armbruster et al. were corroborated by research based on actual student performance data (Eaton & Lovitt, 1972). Eaton and Lovitt reported large inconsistencies in the scores obtained by the same learning disabled students on different achievement tests, and in the degree to which those achievement tests reflected the students' actual performance in particular reading series. On the basis of these studies, one can conclude that standardized achievement tests are unevenly sensitive to student progress and relative standing in varying curricula, and that the use of different tests might result in different educational decisions.

Criterion-Referenced Assessment as an Alternative

In contrast to standardized achievement tests that assess a student's relative standing on global skills, criterion-referenced tests are designed to measure attainment of specific skills within curricula in terms of designated performance standards. Criterion-referenced tests are intended to be an integral part of an instructional system. In principle, the content validity of criterion-referenced tests is strong, since there is close correspondence among a curriculum, objectives, and test items.

The need for criterion-referenced testing has been documented in educational psychology and measurement literature (e.g., Gagne, 1965; Glaser & Nitko, 1971; Popham, 1978). Fifteen years ago Gagne (1965) stated:

> Despite the existence of rather elaborate technology, it cannot be said with confidence that the assessment procedures

> customarily used in developing standardized tests are entirely adequate to meet current assessment needs. One important problem that does not appear to have been included in current techniques is a method for assessing human performance in terms of the objectives of instruction. (p. 258)

Criterion-referenced assessment has received increased attention in the past decade (Popham, 1978). Contributing to this growing popularity have been two related developments. First, educational psychologists have recognized that mastery of subunits precedes mastery of complex tasks. In 1962, Gagne demonstrated the principle that most students can achieve a complex skill providing they have mastered prerequisite component skills. A related development is that educational psychologists (e.g., Bloom, 1971) have applied the above principle in the development of mastery learning systems. In such systems, the curriculum is divided into components and objectives, those objectives are hierarchically arranged, and instruction is directed to the current instructional objective until the student demonstrates mastery of that objective on a criterion-referenced test. Criterion-referenced tests, therefore, are an essential component of a mastery learning system and their use has grown concurrent with the increased popularity of mastery learning systems.

Criterion-referenced assessment of successive units of a curriculum appears to improve upon the content validity of educational assessment. However, in its typical format of pre-post testing around instructional units, criterion-referenced assessment shares a limitation of standardized achievement testing, namely, the poor reliability characteristic of difference scores. Additionally, the time schedule according to which educators administer criterion-referenced tests usually is arbitrary

and most typically determined by the teacher's informal judgment that the child has mastered the skill and is ready to demonstrate this mastery on a test. Consequently, the utility of this measurement format in helping to make decisions about student progress and program adjustments ultimately is dependent upon the accuracy of the teacher's unsystematic monitoring of student performance; that is, how accurately the teacher informally determines that the child is ready to pass the criterion-referenced test. Therefore, criterion-referenced assessment is unsatisfactory for systematically monitoring pupil progress.

Repeated, Curriculum-based Mastery Assessment as an Alternative

Repeated, curriculum-based mastery measurement incorporates the principles of criterion-referenced assessment. It is grounded in the student's curriculum; it measures progress through a hierarchy of objectives in that curriculum; and, it assesses students in relation to performance standards rather than in relation to other students.

However, repeated, curriculum-based mastery measurement departs from the typical criterion-referenced, mastery learning model of assessment in important ways. It borrows the operant research methodology of repeated behavior sampling and time-series analysis. Employing direct and frequent evaluation, a teacher collects repeated, short samples of a student's behavior within the curriculum, over a time period, and under different teaching strategies. At regular intervals, the educator also may measure the performance of mainstream peers on the same behavior. Then, the teacher applies the methods of time-series analysis to the data in order to determine the effectiveness of specific program changes.

Figure 1 illustrates repeated mastery assessment. The abscissa

represents school days and the ordinate represents successive segments or objectives of the curriculum mastered; each data point represents the number of curriculum segments mastered on a given day. The line of best fit through the data points depicts the rate of student progress through the curriculum. The goal of repeated mastery assessment is to increase the student's rate of mastery in the curriculum. The teacher measures the student on a random sample of material from the current instructional curriculum unit until mastery is achieved, at which point (a) the student's graph registers that a curriculum unit has been mastered, (b) the student's level of instruction progresses to the next segment in the hierarchy, and (c) the pool of material on which the teacher measures the student also progresses to the next segment in the hierarchy.

Insert Figure 1 about here

In several ways, direct and repeated mastery measurement appears to strengthen the data base on which educational decisions are made. It improves upon the content validity of norm-referenced tests by evaluating student performance in relation to mainstream functioning on curriculum tasks. It enhances the reliability of measurement because it is administered frequently and therefore is subject to less error and to richer analysis. Furthermore, given the typically short duration of tests and the availability of multiple test forms, it can be employed continuously to evaluate the appropriateness of educational programs.

Nevertheless, mastery measures present two problems. First, they lack the demonstrated construct validity of psychometrically adequate norm-referenced tests. Second, it remains unclear whether performance

on these measures provides information on a student's standing relative to a large, representative normative group. Expressed more concretely, the practitioner's concern is whether a student who manifests progress within mastery measurement (as depicted in Figure 1), can be expected also to show improved performance on traditionally accepted, psychometrically-sound standardized achievement tests.

If simple mastery tests can be shown to demonstrate these characteristics, then direct and repeated measurement might represent a technically adequate educational measurement format that simultaneously provides essential decision-making information. It may, in fact, represent the satisfactory data base with which educators can make and document their decisions.

The purpose of the present investigation was (a) to assess the extent to which simple, direct, progress measures represent the same constructs as longer, more global achievement tests, (b) to determine whether performance on simple tests provides information on students' standing relative to the populations on which norm-referenced achievement tests were standardized, and (c) to investigate whether progress depicted on a mastery graph correlates with progress on psychometrically-sound achievement tests. Reading achievement was selected for the focus of investigation and the study's purpose was translated into three research questions:

- Does performance on simple curriculum-based mastery measures demonstrate concurrent validity with respect to performance on standardized reading achievement tests?

- Is the strength of association between simple curriculum-based mastery and standardized reading achievement tests dependent on the instructional criterion employed?
- Is the strength of association between simple curriculum-based mastery and standardized reading achievement tests dependent on the specific reading material employed?

These research questions deal with the concurrent validity of direct, curriculum-based mastery measures. Concurrent validity studies examine the usefulness of a measure in predicting performance on other variables. Typically, one is interested in assessing the suitability of substituting a short, simple test for a longer and/or more cumbersome criterion that has demonstrated technical adequacy (Messick, 1980). Criterion-relatedness is determined by correlational analysis where the strength of a correlation between two measures specifies the degree of predictive efficiency between the tests (Nunnally, 1967). Therefore, if criterion validity between simple measures and achievement tests is demonstrated and correlations are high, then predictive efficiency would be demonstrated between the tests. On that basis, one might assume that (a) simple tests demonstrate the validity of and represent the same constructs as the longer, more global achievement tests, (b) the simple tests provide information on students' standings relative to the normative population on which the criterion tests were standardized, and (c) as a student manifests improvement on the simple measure, his/her standing relative to the normative group also may improve.

A study of the concurrent validity of simple, direct curriculum-based mastery measures with respect to technically adequate norm-referenced

achievement tests should make an important empirical contribution to the field of measurement and special education. If strong predictive efficiency is demonstrated between these measures, then one might state that simple, direct, and repeated curriculum-based mastery measurement not only encompasses the technical strengths of norm-referenced measurement, but also, as illustrated above, may be more suitable for providing a data base on which educational decisions can be made. Additionally, if strong predictive efficiency is demonstrated between these measures, then the public's acceptance of direct, repeated measurement might be enhanced legitimately to a level at least comparable to that of norm-referenced testing. This has implications for the usefulness of direct, repeated measurement, because the way in which assessment information is accepted may be an important factor in the extent to which, and the ways in which, data are employed and interpreted (Ysseldyke, Algozzine, Regan, Potter, Richey, & Thurlow, 1980).

It also appears important to determine whether the concurrent validity of curriculum-based mastery measures is dependent on the curriculum employed. By definition, direct measurement occurs in the specific curriculum employed within a school. Each curriculum, then, represents a different measure that needs individual validation. This represents a difficult, if not impossible, task. If it can be demonstrated that the specific curriculum employed does not affect the criterion validity of the measure or strength of association between measures, then the need to validate each curriculum separately may be eliminated.

Finally, because progress measurement entails determining mastery

on successive levels of material, it is critical to examine how different performance standards affect the criterion validity of the measure or the strength of association between measures. An examination of the technical adequacy of different performance standards has potential implications for practitioners who employ all formats of criterion-referenced measurement. Criterion-referenced measurement has been criticized repeatedly and severely because of the lack of empirical support for its performance standards (McLoughlin & Lewis, 1981; Thorndike, 1971; Wallace & Larsen, 1978). The present investigation may provide some empirical support for one or more performance standards.

In addition to the three primary research questions, three other related questions were addressed. These questions explored other technical characteristics of repeated, mastery assessment.

The first two questions addressed the congruency of students' instructional scores derived in direct measurement with their performance on more widely accepted criterion measures. These two questions supplemented the second research question, which examined the relationship between the concurrent validity of simple direct measures and the performance standards employed. Because it is possible, theoretically, for two measures to correlate well but agree poorly (Bradley, 1977), when selecting among performance standards, one might consider congruency along with concurrent validity. Specifically, the first related question was: Is the degree of congruency between instructional level scores calculated on curriculum materials and teacher judgments of instructional level scores in the same material dependent on the performance

standard employed? The second related question was: Is the extent of agreement between instructional level grade scores and the achievement test grade scores dependent on the performance standard employed?

The third related question addressed the potential sensitivity of the mastery measures to student achievement. Measures on which students manifest a relatively large range of behavior provide greater opportunity for students to register relatively small gains. A large range of potential behavior that results in heightened sensitivity to student growth is a desirable characteristic of repeated measurement. Therefore, the last related question asked: Is the range of behavior or the average progress per grade level dependent on the performance standard employed?

Method

Subjects

Subjects were 91 randomly selected children, distributed across grades one through six, in one metropolitan public elementary school in the Midwest. All children were English speaking. Fifteen received special education resource service and 23 were enrolled in Elementary and Secondary Education Act Title I programs for children with reading problems.

Measures

Three types of measures were employed in the study: standardized achievement tests, teacher judgments, and graded reading passages.

Standardized achievement tests. The Word Identification (WI) and Passage Comprehension (PC) tests of the Woodcock Reading Mastery Tests (Woodcock, 1973), Form A were employed. The WI test consists

of 150 words ranging in difficulty from preprimer level to words of above average difficulty for twelfth grade students. The easier items were selected from the vocabulary introduced in seven basal reading programs from the first preprimer through the third grade reader (Woodcock, 1973). The more difficult items were drawn primarily from the Thorndike-Lorge List (Thorndike & Lorge, 1944). The subject's task in the WI test is to name words.

The PC Test contains 85 items of a modified cloze procedure (Bormuth, 1969). The subject's task is to read silently a passage from which a word has been deleted and to supply orally to the examiner an appropriate missing word. The passages range in difficulty from first grade to college level (Woodcock, 1973).

Teacher judgments. For each student, teachers reported the book level in Ginn 720 (1976) from which the student read for instruction.

Reading passages. Reading passages from the Ginn 720 and the Scott-Foresman Unlimited (1976) series were employed in measurement. For 10 levels in Ginn and 9 levels in Scott-Foresman, two 100-word reading passages were selected as representative of the average readability level of the material from which the passages were drawn. Representative passages were employed because of Fitzgerald's (1980) finding of great variability in the readability of series of passages from the same books within seven reading series. Within repeated measurement, the effect of this variability on the reliability of a student's score is diminished because an average or median level across multiple observations is employed to describe a pupil's performance. In the current study, it was not feasible to measure students repeatedly.

So, in an attempt to improve the reliability of students' scores to represent more accurately the technical adequacy of repeated measurement, passages were selected as representative of the average readability of the material from which they were drawn. (See Fuchs & Deno, 1981, for a description of the passage selection procedure.) Table 1 displays publishers' level numbers and grade levels, and readability information for each selected passage of both series.

Insert Table 1 about here

Procedure

Prior to testing, the classroom teachers completed and returned to the investigator a form on which they indicated the students' actual Ginn placements. Also, five examiners were identified and trained in the administration and scoring of all measures.

During a 45 to 60 minute session, each subject was tested individually on all measures, by one randomly determined examiner in one of four quiet and isolated locations within the school. The WI and PC tests were administered according to the Manual (Woodcock, 1973). For both series, the reading passages were administered in a random order employing the following procedure: The examiner found the appropriate passage in a teacher notebook containing all passages and the corresponding passage in a student notebook containing all passages. As the examiner exposed the passage to the student, the examiner said, "I'd like you to read aloud some words to me as quickly as you can. If you don't know a word, skip it. Try your hardest. Remember to read very quickly. I'll tell you when to stop. Any questions?" The examiner said "Begin" as he/she

started a stopwatch; as the student read, the examiner wrote with a transparency pen on the acetate covering the teacher copy. Making sure that his/her writing was hidden from the subject, the examiner crossed out omissions, substitutions, insertions, and mispronunciations. If the student completed a passage in less than 60 seconds, the examiner wrote the number of seconds in which the student read the passage. At 60 seconds, the examiner told the student to stop. With each passage, the examiner repeated this procedure except for directions, where the examiner simply said, "Any questions? Ready to read?" After all testing was completed for a student, the examiner scored each passage by counting words correct and words incorrect. Then on a recording form, the examiner wrote these scores in the appropriate spaces and indicated the number of seconds for those passages that the student completed in less than 60 seconds.

For each of the 19 passages, each student received a words correct per minute score, an errors per minute score, and a percent correct per minute score. On the basis of these scores, a student was assigned, within each series, seven different instructional level scores based on the following criteria of instructional level:

> Instructional Criterion 1: For preprimer (PP) through grade 3 books, 30-49 words per minute (wpm) with 7 or fewer errors per minute (epm); for grade 4 through grade 6 books, 50+ wpm with 7 or fewer epm (Starlin & Starlin, 1974).
>
> Instructional Criterion 2: 70+ wpm with 10 or fewer epm (Starlin, 1979).
>
> Instructional Criterion 3: 100+ wpm with 0-2 epm (Haring, Liberty, & White, undated).
>
> Instructional Criterion 4: 95% accuracy (Betts, 1946; Harris, 1961; Powell, 1971).

Instructional Criterion 5: 70+ wpm and 95% accuracy.

Instructional Criterion 6: For PP through grade 2 books, 50+ wpm and 95% accuracy; for grades 3 through 6, 70+ wpm and 95% accuracy.

Instructional Criterion 7: For PP through grade 2 books, 50+ wpm and 85% accuracy (Powell, 1971); for grades 3 through grade 6 books, 70+ wpm with 95% accuracy.

For each criterion, within each series, an instructional level score was assigned to each student by identifying the highest level at which the criterion was met before an unsatisfactory performance was demonstrated at two consecutive levels.

Results

Do Simple Curriculum-based Mastery Measures Predict Performance on Standardized Reading Achievement Tests of Word Identification and Comprehension?

To examine this question, a Pearson Product-Moment correlation matrix was generated, including the seven instructional level scores for both series and the PC and WI raw scores. Table 2 displays these correlations. Inspection of this table reveals that correlations were moderate to high and were statistically significant ($p < .001$). The correlations ranged from .57 to .95; 23 of the 28 correlations were greater than or equal to .80.

Insert Table 2 about here

Is the Strength of Association Between a Simple Curriculum-based Mastery Measure and a Standardized Reading Achievement Test Dependent on the Instructional Criterion Employed?

To investigate this question, the correlations in Table 2 were examined across series and within instructional criteria. Averaged

within instructional criteria, at least one criterion of instruction appeared to affect the strength of association between the simple curriculum-based mastery measures and the standardized reading achievement tests. For Criterion 3, the average of the four correlations was .62, lower than any other average correlation by .23. The average correlation was highest for Criterion 1 (.93). The correlations produced by the remaining criteria were similar, ranging from an average .85 for Criteria 4 and 5 to an average .87 for Criteria 2 and 7.

Unfortunately, there is no appropriate statistical test for determining the difference between two dependent correlations when one of the sets of scores is not identical in both correlations. This makes it impossible to test statistically the difference between many of the correlations calculated on the same sample in this study. Additionally, where questions concern differences between two dependent correlations when one of the sets of scores is identical in both correlations (r_{xz} with r_{yz}), the available test limits any inference to only a subpopulation of all possible samples for which X and Y have exactly the same set of values as those in the observed sample (Walker & Lev, 1969). Consequently, the utility of such a test is limited, and given the dependency in the data and the large number of analyses run, it appears appropriate to forego these additional statistical analyses (Terwilliger, 1980). Therefore, the differences in the data are discussed without the benefit of statistical probability.

Is the Strength of Association Between a Simple Curriculum-based Mastery Measure and a Standardized Reading Achievement Test Dependent on the Specific Reading Material Employed?

For each instructional criterion, the correlations in Table 2 were examined within series. As is evident in the table, all correlations within series were high and statistically significant ($p < .001$). For Series A the mean correlation (.87) was somewhat higher than for Series B (.82). Furthermore, within each instructional criterion and within achievement test (PC vs WI), the correlation for Series A was consistently higher than for Series B. However, the difference between these mean correlations (.87 - .82 = .05) was small and probably does not represent a reliable difference.

Additional Analyses

In addition to the research questions explored above, three analyses relating to other technical characteristics of repeated mastery measurement were completed on the data collected in this study.

Congruency between instructional level scores calculated on curriculum materials and teacher judgments of instructional level scores in the same material and relationship of congruency to the performance standard employed. The degree of congruency between instructional scores and teacher placements was examined by calculating, for each instructional criterion, the percentages of students whose instructional level scores placed them the same as, below, or above the teacher placements. These percentages are displayed in Table 3. Inspection of this table reveals that Instructional Criteria 4, 5, 6, and 7 were similar in congruency, with an average 19.5% of students placed below, 64.5% of students placed the same, and an average of 15.8% of students placed above the

teacher placement. The distribution for Criterion 2 was similar to Criteria 4 through 7; however, it placed a greater percentage of students (29.0%) above the teacher placement. Criterion 1 placed a great percentage of students (50.0%) above while Criterion 3 placed the greatest percentage of students (58.0%) below the teacher placement.

Insert Table 3 about here

Correlated *t* tests corroborated this pattern of congruency for the different instructional criteria. The difference between the instructional scores and the teacher placement was statistically significant for Criterion 1, $t(89) = 8.42$, $p = .000$ (mean difference = 1.87), and for Criterion 2, $t(89) = 2.29$, $p = .000$ (mean difference = .54). For Criterion 3, the difference also was statistically significant $t(89) = -7.72$, $p = .000$. For this criterion, however, the teacher placements were above the instructional scores (mean difference = -2.32). For Criteria 4 through 7, there was no statistically significant difference.

Agreement between the instructional grade scores and the achievement test grade scores and relationship to the criterion of instructional level employed. The degree of congruency between instructional grade scores and achievement test grade scores was examined by calculating, for each instructional criterion, the percentages of students whose instructional grade scores placed them below, at the same level, and above the PC and WI grade scores. Therefore, four combinations of congruency percentages were calculated: Series A instructional grade

scores with PC and with WI grade scores, and Series B instructional grade scores with PC and with WI grade scores. The average percentages across the four combinations are presented in Table 4.

Insert Table 4 about here

The extent of congruency was similar again for Criteria 4, 5, 6, and 7, with averages across the four criteria being 51.39% of students placed the same, 10.18% placed above, and 38.43% placed below the achievement grade scores. Criterion 2 presented a similar pattern with approximately equal percentages placed below and above the achievement scores. Criterion 3 placed a great percentage of students (60.25%) below, while Criterion 1 placed a great percentage of students (43.25%) above.

Again, correlated t tests corroborated this pattern of congruency for the instructional criteria. For Criteria 1 and 3, the difference between the instructional grade scores and the achievement test grade scores was always statistically significant ($t(91) \geq 3.35$, $p < .001$ for Criterion 1, and $t(91) \geq 5.33$, $p = .000$ for Criterion 3). Criterion 1 placed students above by an average of .55 levels; Criterion 3 placed students below by an average of 1.29 levels. For Criterion 2, the average difference was the smallest (.11 levels).

Average increase per grade level as a function of the instructional criterion employed. Within series and for each instructional criterion, the mean instructional level score for each grade level was graphed (see Figures 2 and 3). Next, by series and by instructional criteria, the average increase per grade level was calculated. Finally, across series, these means were averaged (see Table 5).

Insert Table 5 and Figures 2 and 3 about here

Visual inspection of Figures 2 and 3 and analysis of Table 5 reveals that, across series, the average increase per grade level was similar for all criteria but the third, where the average increase was relatively small.

Discussion

In analyzing the success of the criterion-based measures in predicting achievement test performance, the statistically significant, moderate to high correlations provided evidence for the concurrent validity of curriculum-based mastery reading measures with word recognition and comprehension achievement tests. In a comparison between correlations with Passage Comprehension and with Word Identification scores, correlations within instructional criteria and series were always similar, even though criteria did not require students to demonstrate any comprehension of the material. This may be explained by the fact that the Woodcock Passage Comprehension Test uses a cloze procedure that asks students to read words rather than to answer comprehension questions. Nevertheless, the progress measures do appear to predict performance on both valid and reliable standardized tests of reading comprehension and words.

Seven instructional criteria based on oral reading in context were employed to explore the dependence of the above association on the performance standard employed. Criteria 1 through 3 were selected because they are advocated by Precision Teachers (Alper, Nowlin, Lemoine, Perine, & Bettencount, 1973; Haughton, 1972; Starlin, 1979; Starlin & Starlin, 1974). Criterion 4 was employed because it is the traditionally accepted

informal reading inventory, instructional criterion of word recognition accuracy (Betts, 1946; Harris, 1961; Powell, 1971). Criteria 5 and 6 represented combinations of the rate and percentage-accuracy criteria found in the first three criteria. In Criterion 7, a lower standard of 85% accuracy for students in books preprimer through grade 2 was introduced because Powell (1971) demonstrated that preprimer through second grade readers maintained 70% comprehension while their word recognition accuracy was at 85% or better.

All correlations between the instructional scores and the Passage Comprehension and Word Identification scores were moderate to high and statistically significant regardless of the instructional criterion employed. Yet, a careful comparison among the average correlations associated with each instructional criterion revealed that at least one criterion of instructional level was a differentially poor predictor. Criterion 3, the most stringent criterion, placed many students at low reading levels, failing to discriminate effectively among readers with different skills, resulting in lower correlations with achievement tests, and failing to predict efficiently performance on achievement tests. Therefore, the strength of association between curriculum-based mastery measures and standardized reading achievement tests does appear to be affected by the instructional criterion employed. Results of this study suggest that as practitioners select an instructional criterion to employ within direct and repeated curriculum-based measurement, they might opt for rates between 30 and 70, and/or percentages between 85 and 95.

In contradistinction to these results, Beck (1980), in the Sacajawea

Project, suggested rates as high as 150 wpm. The discrepancy between Beck's recommendations and those based on this study may be explained by a distinction between proficiency and fluency (Brent, Arnold, & DuRoss, 1978). "Proficiency" is the level of performance standard that results in long-term maintenance from intense practice. "Fluency" is that level of performance that represents competency on unfamiliar material. In this study, fluency was assessed. Children read primarily unfamiliar material; from that performance, judgments of instructional placements were determined. However, Beck's interest is proficiency, or the level at which, after intense practice, a student can progress to new material. It may be that, with familiar material, the more stringent mastery criterion of 150 wpm might result in higher correlations and in better congruency.

The third major question addressed in the present study asked whether the association between curriculum-based measures and standardized achievement tests was dependent on the reading curriculum employed. To explore this question, two basal reading series were selected that had different program emphases. One of the series employed, Ginn 720, is representative of many basal programs in its eclectic approach to reading instruction. The other program, Scott-Foresman Unlimited series, places a greater emphasis on comprehension and study skills. It was reasoned that if the strength of association demonstrated by two different types of reading curricula were similar, then one might generalize that the strength of association would be similar across other curricula as well.

Across criteria and within the two series, all correlations were statistically significant, high, and similar. Neither the criterion

validity of these measures, then, nor the strength of their association with standardized reading achievement tests appears to be dependent on the reading material employed. To the extent that the curricula used are representative of basal reading series, then curriculum-based mastery measurement from different reading curricula demonstrates strong predictive efficiency and concurrent validity with respect to achievement tests. Apparently, the practitioner might assume that the selection of reading series does not affect the validity of curriculum-based measurement and that curriculum-based measures can be used across different basal reading series.

In addition to the major research questions of this study, the issue of congruency between curriculum-based measurement and both teacher judgments and achievement tests was explored. These two analyses were conducted because it is possible, theoretically, for two measures to correlate well but agree poorly (Bradley, 1977). In selecting among instructional criteria, one might well consider congruency along with the strength of association.

In the present study, the results revealed that, first, the degree of congruency between teacher placements and the curriculum-based placements varied with the instructional criteria used. Second, the extent of agreement between curriculum-based mastery measures and achievement test grade scores was different when different instructional criteria were employed. The degree of criterion validity of curriculum-based measures appeared to be dependent on the criteria employed in the measurement. Additionally, the results empirically demonstrated Bradley's

contention that it is possible for two measures to correlate well but agree poorly. For example, Criterion 1 produced the highest average correlation but did not agree well with either of the criterion measures.

As the practitioner selects a criterion of instructional level to implement within repeated curriculum-based mastery measurement, he/she might opt for one of the other instructional criteria that simultaneously produced good correlations and agreed well with the criterion measures. Several legitimate standards might be selected for determining an acceptable mastery criterion. Assuming as a dual standard for good agreement at least 50% equivalent placements and an educationally unimportant difference of .50 level or less between mastery level (grade) scores and other criteria, Instructional Criteria 2, 4, 6, and 7 appear acceptable. Criterion 2 is 70+ wpm with 10 or fewer errors across grade levels. Criterion 4 is 95% accuracy (in a one-minute sample). Criteria 6 and 7 employ different oral reading rates for primary (50 wpm) and intermediate (70 wpm) readers, with 95%/95% or 85%/95% accuracy criteria, respectively. Any one of these four criteria appears to be a good choice for practitioners. Additionally, for ongoing use of a mastery criterion where one is interested in proficiency rather than fluency, one might consider Beck's recommendation of 150 wpm. However, the external validity of this criterion is unclear.

When selecting between percentage and percentage-rate criteria, there are several instructional considerations. Precision Teaching experts (Cohen, 1975; Lindsley, 1971; Haughton, 1969) argue that rate (a) is more sensitive to behavioral change than is percentage, (b) provides a basis for comparing performance among curricula, (c) communicates speed and accuracy rather than just accuracy, and (d) imposes no performance

ceiling as does percentage. Further, although percentage implies a reciprocal relationship between correct and incorrect responses, this is not necessarily the case. It is possible for a student to score 90% on two days and for that student's performance to be qualitatively different. The same percentage score may be based on differing numbers of errors and words correct from day to day. Therefore, the combination of rate and percentage communicates more information than percentage alone; for instructional planning purposes, the practitioner may prefer one of the percentage rate combination performance standards. At the same time, it should be remembered that in the present study the reading sample was time limited, so even the 95% criterion was in some sense a ratio and accuracy criterion.

Another related issue addressed in the present study was the potential sensitivity of different measures to student achievement progress. Measures on which students manifest a relatively large range of behavior provide greater opportunity for students to register relatively small gains. A large range of potential behavior that results in heightened sensitivity to student growth is a desirable characteristic of repeated measurement.

Criterion 3 appeared to produce a differentially low rate of average progress both within and across series, suggesting that the average progress per grade level might be dependent on the instructional criterion employed. Yet, across the six remaining criteria, there appeared to be no effect. This leads one to infer that only the third, most stringent criterion, which also resulted in relatively poor association and poor agreement with criterion measures, differentially affected the average progress per grade.

Conclusions

The research findings in this investigation support several generalizations. They are summarized as follows:

- The validity of simple curriculum-based mastery measurement is strong. Performance on curriculum-based mastery reading measures is highly related to performance on a valid and reliable standardized reading achievement test.
- The validity of simple curriculum-based mastery measures is dependent on the instructional criterion employed.
- The validity of simple curriculum-based reading measures appears to be independent of the specific reading series employed.
- The degree of congruency between instructional level scores calculated by simple curriculum-based measures and teacher placements is dependent on the instructional criterion employed.
- The extent of agreement between the instructional grade scores calculated by simple curriculum-based measures and achievement test grade scores is dependent on the instructional criterion employed.
- In mastery measurement, the range of behavior or average increase per grade level appears to be dependent on the instructional criterion employed.
- Of the instructional criteria employed in this investigation, those employing (a) 70 wpm with 88% accuracy, (b) 95% accuracy, and (c) different oral reading rates for primary (50 wpm) and intermediate (70 wpm) readers with 95%/95% and 85%/95% accuracy,

respectively, are good choices. They render scores that demonstrate strong criterion validity with respect to achievement tests, and they produce scores that agree well with teacher placements and with achievement test scores.

As discussed in the introduction to this paper, simple curriculum-based mastery measures validated in this investigation when employed repeatedly and analyzed with time-series methodology, yield more useful data than traditional testing formats for making educational decisions. First, direct measures evaluate student performance in relation to mainstream functioning. Second, they can be administered frequently and therefore reduce error, enable richer analyses, and allow programs to be evaluated continuously. Finally, because they are direct, their content validity is high, providing useful data concerning student progress in the curriculum where the student functions.

Norm-referenced achievement tests, however, do present three distinct advantages over direct and repeated measurement. They have demonstrated construct validity. They provide information on students' standings relative to large, well-represented groups of children, and they are better accepted among professionals and parents.

Results from this study indicate that simple curriculum-based measures represent the same constructs as the longer, more global tests, and that they can be used to provide information on students' standings not only relative to mainstream peers but also relative to the same large representative population on which the Woodcock Reading Mastery Tests were normed. It also appears that improved performance on direct and repeated measures might indicate improved standing relative to that

large representative population. Therefore, it appears that repeated curriculum-based mastery measurement may fill a void in educational measurement.

References

Alper, T., Nowlin, L., Lemoine, K., Perine, M., & Bettencount, B. The rated assessment of academic skills. Academic Therapy, 1973, 9, 151-164.

Anastasi, A. Psychological testing (4th ed.). New York: Macmillan, 1976.

Armbruster, B. B., Stevens, R. J., & Rosenshine, B. Analyzing content coverage and emphases: A study of three curricula and two tests (Technical Report No. 26). Urbana-Champaign: University of Illinois, Center for the Study of Reading, 1977.

Beck, R. Personal communication, July, 1980.

Betts, E. A. Foundations of reading instruction. New York: American Book, 1946.

Bloom, B. S. Learning for mastery. In B. S. Bloom, J. T. Hasting, & G. F. Madaus (Eds.), Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill, 1971.

Bormuth, J. R. Factor validity of cloze tests as measures of reading comprehension ability. Reading Research Quarterly, 1969, 4, 358-365.

Bradley, J. M. Evaluating reading achievement for placement in special education. Journal of Special Education, 1977, 10(2), 168-177.

Brent, G., Arnold, H., & DuRoss, B. Urban applications of precision teaching to reading. Reading Improvement, 1978, 15(3), 190-202.

Cohen, M. Teacher judgment concerning arithmetic disabilities accounted for by a functional assessment battery. Unpublished doctoral dissertation, University of Washington, 1975.

Eaton, M., & Lovitt, T. Achievement tests vs. direct and daily measurement. In G. Semb (Ed.), Behavior analysis and education. Lawrence, Kan.: University of Kansas, 1972.

Fitzgerald, G. G. Reliability of the Fry sampling procedure. Reading Research Quarterly, 1980, 15(4), 489-503.

Fuchs, L. S., & Deno, S. L. A comparison of reading placements based on teacher judgment, standardized testing, and curriculum-based assessment (Research Report No. 56). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1981.

Gagne, R. The acquisition of knowledge. Psychological Review, 1962, 69, 355-365.

Gagne, R. The conditions of learning. New York: Holt, Rinehart, & Winston, 1965.

Ginn and Company. Reading 720. Lexington, Mass.: Ginn (Xerox Corp.), 1976.

Glaser, R., & Nitko, J. Measurement in learning and instruction. In R. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Harris, A. J. How to increase reading ability. New York: David McKay, 1961.

Haring, N., Liberty, K., & White, O. Instructional hierarchies research project. Unpublished manuscript. (University of Washington, Seattle)

Haughton, E. Aims - growing and sharing. In J. Jordan & L. Robbins (Eds.), Let's try doing something else kind of thing. Arlington, Va.: The Council for Exceptional Children, 1972.

Jenkins, J., Deno, S., & Mirkin, P. Measuring pupil progress toward the least restrictive environment (Monograph No. 10). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979.

Jenkins, J., & Pany, D. Standardized achievement tests: How useful for special education? Exceptional Children, 1978, 44(6), 448-453.

Lindsley, O. R. Operant conditioning methods applied to research in chronic schizophrenia. Psychiatric Research Reports, 1956, 5, 118-153.

McLoughlin, J., & Lewis, R. Assessing special students: Strategies and procedures. Columbus, Ohio: Charles E. Merrill, 1981.

Messick, S. Test validity and the ethics of assessment. American Psychologist, 1980, 35(11), 1012-1027.

Nunnally, J. C. Psychometric theory. New York: McGraw-Hill, 1967.

Popham, W. J. Criterion-referenced measurement. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Powell, W. K. Validity of the IRI reading levels. Elementary English, 1971, 48, 637-642.

Scott-Foresman Systems, Revised. Unlimited Series. Glenview, Ill.: Scott-Foresman, 1976.

Stanley, J. C. Reliability. In R. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Starlin, C. Evaluating and teaching reading to "irregular" kids. Iowa Perspective, Iowa Department of Public Instruction, Dec., 1979, 1-11.

Starlin, C., & Starlin, A. Guidelines for continuous decision making. Bemidji, Minn.: Unique Curriculums Unlimited, 1974.

Terwilliger, J. Personal communication, 1980.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Teachers College, Columbia University, 1944.

Thorndike, R. Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Tyler, R. W. The functions of measurement in improving instruction. In E. F. Lindquist (Ed.), Educational measurement. Washington, D.C.: American Council on Education, 1951.

Wallace, G., & Larsen, S. C. Educational assessment of learning problems: Testing for teaching. Boston: Allyn & Bacon, 1978.

Walker, H., & Lev, J. Statistical inference. New York: Holt, Rinehart, & Winston, 1969.

Woodcock, R. Woodcock Reading Mastery Tests Manual. Circle Pines, Minn.: American Guidance Service, 1973.

Ysseldyke, J. Psychoeducational assessment and decision making. In J. E. Ysseldyke & P. Mirkin (Eds.), Proceedings of the Minnesota roundtable conference on assessment of learning disabled children (Monograph No. 8). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979.

Ysseldyke, J., Algozzine, B., Regan, R. R., Potter, M., Richey, L., & Thurlow, M. L. Psychoeducational assessment and decision making: A computer simulated investigation (Research Report No. 32). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1980.

Table 1

Level Numbers, Grade Levels, and Readability

Information of Passages from Two Reading Series

Series[a] Level Number	Grade Levels	$\bar{X}$ Readability Score Across Passage	N[b]	SD[c]	$\bar{X}$ Readability Scores of Two Selected Passages
Series A					
3-4	PP-P	2.02	8	.098	2.01
5	1-1	2.21	5	.117	2.20
6	2-1	2.43	6	.196	2.43
7	2-2	3.17	13	.536	3.10
8	3-1	3.60	10	.468	3.66
9	3-2	4.11	6	.142	4.05
10	4	5.00	11	.476	5.00
11	5	5.38	10	.534	5.36
12	6	5.81	14	.392	5.75
13	7	6.00	13	.593	6.03
Series B					
2-3	PP-P	2.57	9	.439	2.57
4	1	2.73	5	.156	2.77
5-6	2-1	2.87	10	.282	2.95
7-8	2-2	3.29	7	.293	3.30
9-10	3-1	3.64	9	.754	3.59
11-12	3-2	4.02	13	.520	3.94
13-15	4	4.89	5	.252	4.82
16-18	5	5.64	11	.525	5.70
19-21	6	6.04	13	.144	6.03

[a]Series A is Ginn 720 and Series B is Scott-Foresman Unlimited.

[b]Number of passages employed.

[c]Standard deviation across passages.

Table 2

Correlations Between Instructional Scores on Simple Measures and Raw Scores on the Passage Comprehension (PC) and Word Identification (WI) Achievement Tests (N=91)

Instructional Criterion	Series[b]	Correlation[a] with PC	Correlation[a] with WI
1	A	.93	.95
1	B	.92	.92
2	A	.92	.89
2	B	.87	.82
3	A	.65	.62
3	B	.63	.57
4	A	.88	.88
4	B	.82	.81
5	A	.90	.88
5	B	.83	.78
6	A	.91	.89
6	B	.85	.80
7	A	.93	.91
7	B	.89	.86

[a]All correlations were statistically significant ($p < .001$).

[b]Series A is Ginn 720 and Series B is Scott-Foresman Unlimited.

Table 3

Percentages of Students Placed Below, the Same, and Above Teacher Placements by Each Instructional Criterion (N=89)[a]

Instructional Criterion	Placement by Curriculum-based Measures Compared to Teacher Placement		
	Below	Same	Above
1	3	47	50
2	18	53	29
3	58	39	3
4	21	61	18
5	23	63	15
6	19	65	14
7	15	69	16

[a]No placement was reported for two students.

Table 4

For Each Instructional Criterion, Percentages of Students Placed Below, the Same, and Above Achievement Test Scores (N=91)[a]

Instructional Criterion	Curriculum-based Grade Scores Compared to Achievement Test Scores		
	Below	Same	Above
1	11.25	44.75	43.25
2	26.50	51.50	21.50
3	60.25	38.00	1.00
4	39.25	46.50	13.50
5	42.50	49.00	7.75
6	40.00	51.75	7.50
7	32.50	58.00	8.75

[a]Percentages are across reading series and across achievement tests (WI and PC).

Table 5

Average Increase Per Grade Level Calculated on Curriculum-based Measures

Instructional Criterion	Average Increase Per Grade Level		
	Series A	Series B	Average Across Series
1	1.90	1.76	1.83
2	1.90	1.72	1.81
3	.96	.94	.95
4	1.76	1.74	1.75
5	1.76	1.68	1.72
6	1.88	1.64	1.76
7	1.88	1.68	1.78

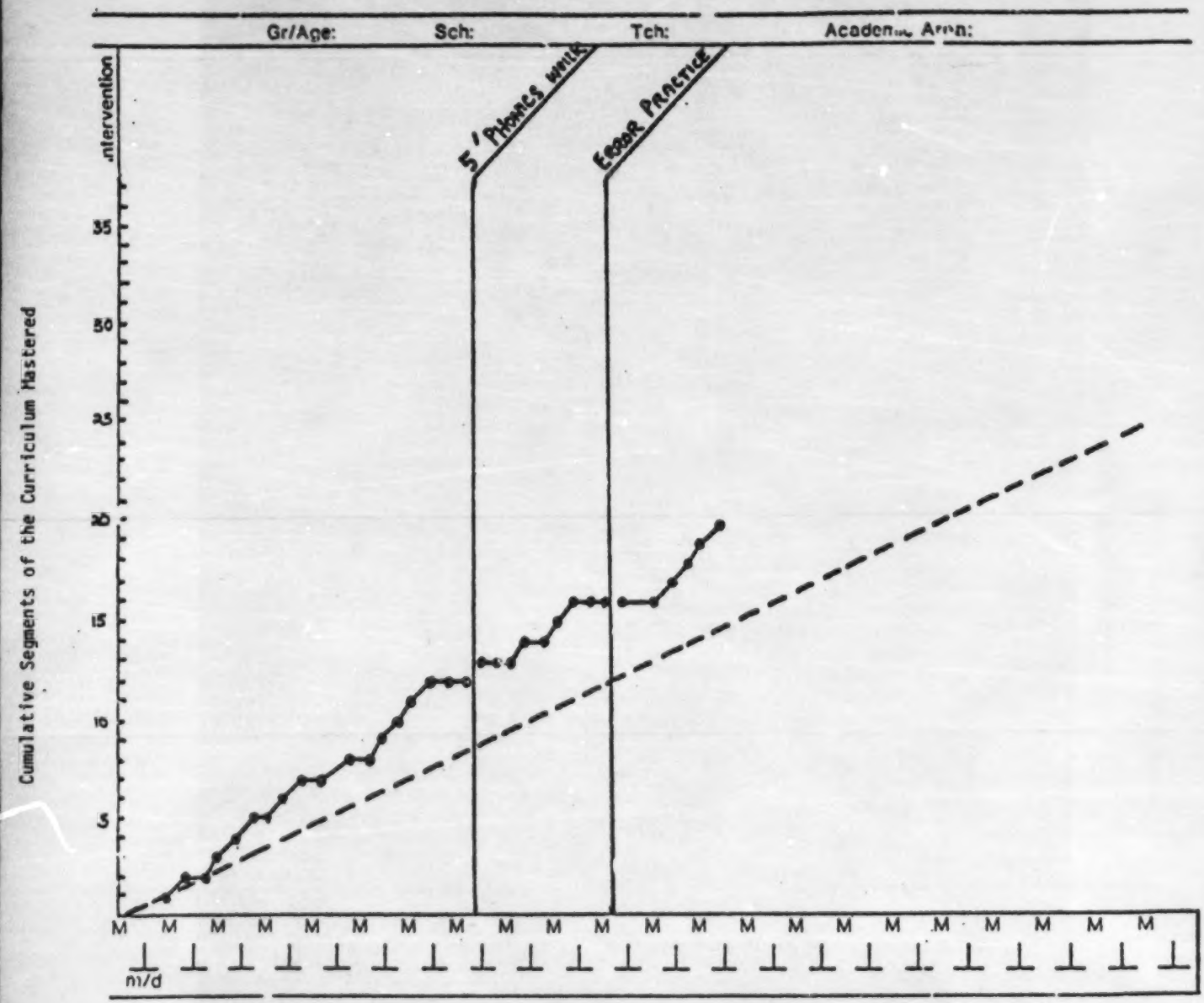

Figure 1. Illustration of Repeated Mastery Measurement.

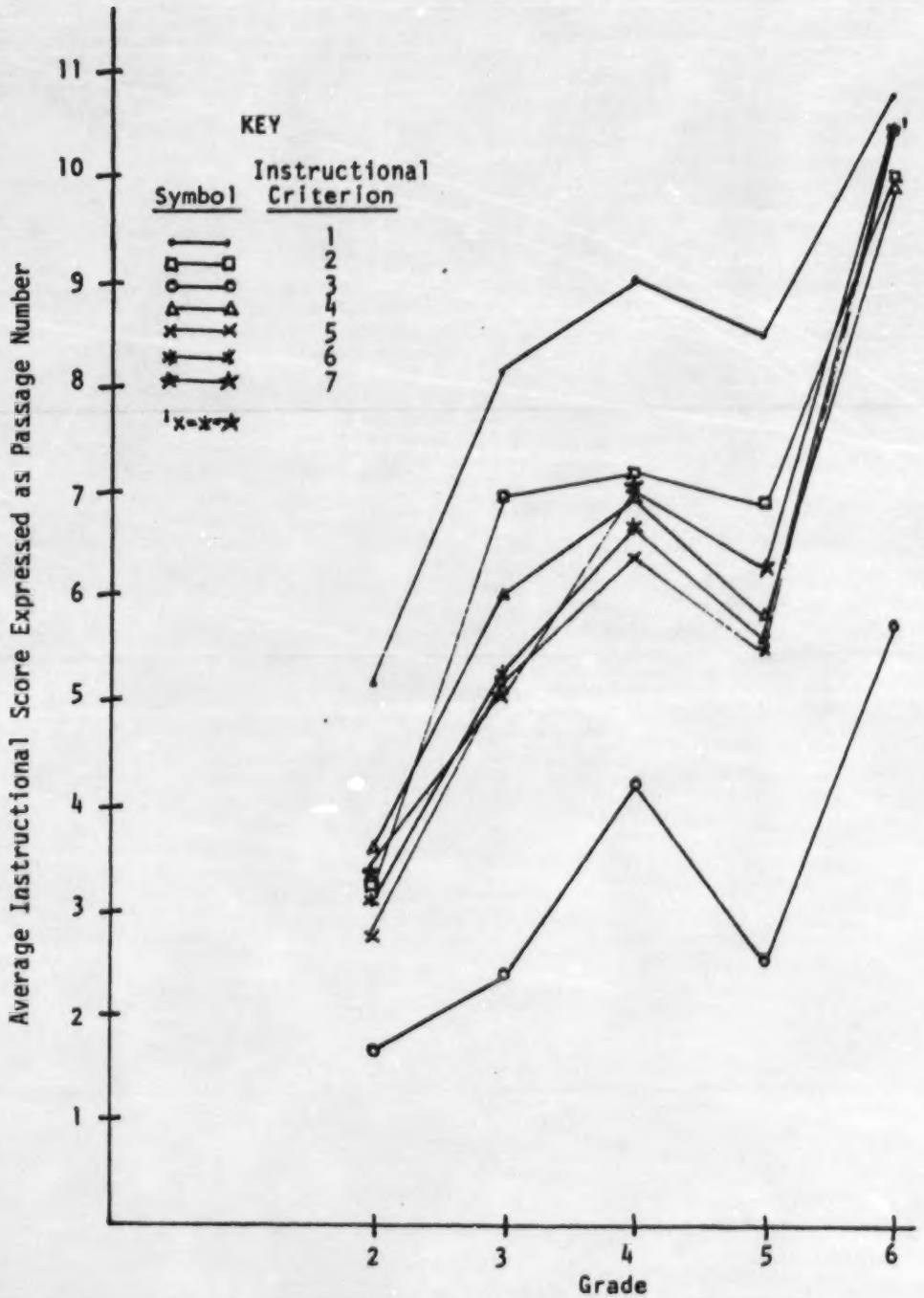

Figure 2. Within the Ginn 720 series for each instructional criterion, the average instructional score per grade level.

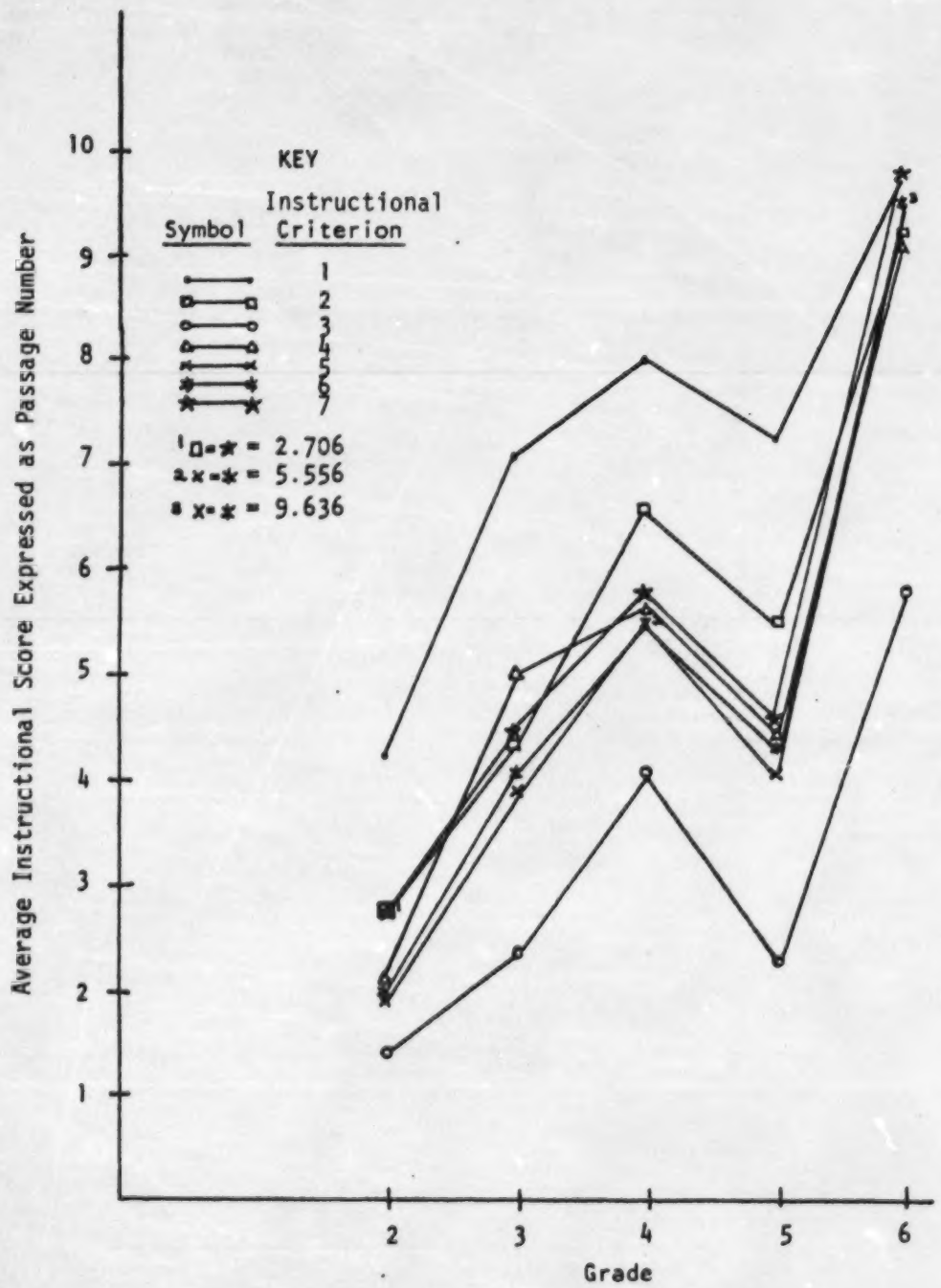

Figure 3. Within the Scott-Foresman Unlimited series, for each instructional criterion, the average instructional score per grade level.

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $3.00 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.

Ysseldyke, J. E. Assessing the learning disabled youngster: The state of the art (Research Report No. 1). November, 1977.

Ysseldyke, J. E., & Regan, R. R. Nondiscriminatory assessment and decision making (Monograph No. 7). February, 1979.

Foster, G., Algozzine, B., & Ysseldyke, J. Susceptibility to stereotypic bias (Research Report No. 3). March, 1979.

Algozzine, B. An analysis of the disturbingness and acceptability of behaviors as a function of diagnostic label (Research Report No. 4). March, 1979.

Algozzine, B., & McGraw, K. Diagnostic testing in mathematics: An extension of the P:AT? (Research Report No. 5). March, 1979.

Deno, S. L. A direct observation approach to measuring classroom behavior: Procedures and application (Research Report No. 6). April, 1979.

Ysseldyke, J. E., & Mirkin, P. K. Proceedings of the Minnesota roundtable conference on assessment of learning disabled children (Monograph No. 8). April, 1979.

Somwaru, J. P. A new approach to the assessment of learning disabilities (Monograph No. 9). April, 1979.

Algozzine, B., Forgnone, C., Mercer, C. D., & Trifiletti, J. J. Toward defining discrepancies for specific learning disabilities: An analysis and alternatives (Research Report No. 7). June, 1979.

Algozzine, B. The disturbing child: A validation report (Research Report No. 8). June, 1979.

Note: Monographs No. 1 - 6 and Research Report No. 2 are not available for distribution. These documents were part of the Institute's 1979-1980 continuation proposal, and/or are out of print.

Ysseldyke, J. E., Algozzine, B., Regan, R., & Potter, M. Technical adequacy of tests used by professionals in simulated decision making (Research Report No. 9). July, 1979.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. Measuring pupil progress toward the least restrictive environment (Monograph No. 10). August, 1979.

Mirkin, P. K., & Deno, S. L. Formative evaluation in the classroom: An approach to improving instruction (Research Report No. 10). August, 1979.

Thurlow, M. L., & Ysseldyke, J. E. Current assessment and decision-making practices in model programs for the learning disabled (Research Report No. 11). August, 1979.

Deno, S. L., Chiang, B., Tindal, G., & Blackburn, M. Experimental analysis of program components: An approach to research in CSDC's (Research Report No. 12). August, 1979.

Ysseldyke, J. E., Algozzine, B., Shinn, M., & McGue, M. Similarities and differences between underachievers and students labeled learning disabled: Identical twins with different mothers (Research Report No. 13). September, 1979.

Ysseldyke, J., & Algozzine, R. Perspectives on assessment of learning disabled students (Monograph No. 11). October, 1979.

Poland, S. F., Ysseldyke, J. E., Thurlow, M. L., & Mirkin, P. K. Current assessment and decision-making practices in school settings as reported by directors of special education (Research Report No. 14). November, 1979.

McGue, M., Shinn, M., & Ysseldyke, J. Validity of the Woodcock-Johnson psycho-educational battery with learning disabled students (Research Report No. 15). November, 1979.

Deno, S., Mirkin, P., & Shinn, M. Behavioral perspectives on the assessment of learning disabled children (Monograph No. 12). November, 1979.

Sutherland, J. H., Algozzine, B., Ysseldyke, J. E., & Young, S. What can I say after I say LD? (Research Report No. 16). December, 1979.

Deno, S. L., & Mirkin, P. K. Data-based IEP development: An approach to substantive compliance (Monograph No. 13). December, 1979.

Ysseldyke, J., Algozzine, B., Regan, R., & McGue, M. The influence of test scores and naturally-occurring pupil characteristics on psycho-educational decision making with children (Research Report No. 17). December, 1979.

Algozzine, B., & Ysseldyke, J. E. Decision makers' prediction of students' academic difficulties as a function of referral information (Research Report No. 18). December, 1979.

Ysseldyke, J. E., & Algozzine, B. Diagnostic classification decisions as a function of referral information (Research Report No. 19). January, 1980.

Deno, S. L., Mirkin, P. K., Chiang, B., & Lowry, L. Relationships among simple measures of reading and performance on standardized achievement tests (Research Report No. 20). January, 1980.

Deno, S. L., Mirkin, P. K., Lowry, L., & Kuehnle, K. Relationships among simple measures of spelling and performance on standardized achievement tests (Research Report No. 21). January, 1980.

Deno, S. L., Mirkin, P. K., & Marston, D. Relationships among simple measures of written expression and performance on standardized achievement tests (Research Report No. 22). January, 1980.

Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehnle, K. Formative evaluation: Continued development of data utilization systems (Research Report No. 23). January, 1980.

Deno, S. L., Mirkin, P. K., Robinson, S., & Evans, P. Relationships among classroom observations of social adjustment and sociometric rating scales (Research Report No. 24). January, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Factors influential on the psychoeducational decisions reached by teams of educators (Research Report No. 25). February, 1980.

Ysseldyke, J. E., & Algozzine, B. Diagnostic decision making in individuals susceptible to biasing information presented in the referral case folder (Research Report No. 26). March, 1980.

Thurlow, M. L., & Greener, J. W. Preliminary evidence on information considered useful in instructional planning (Research Report No. 27). March, 1980.

Ysseldyke, J. E., Regan, R. R., & Schwartz, S. Z. The use of technically adequate tests in psychoeducational decision making (Research Report No. 28). April, 1980.

Richey, L., Potter, M., & Ysseldyke, J. Teachers' expectations for the siblings of learning disabled and non-learning disabled students: A pilot study (Research Report No. 29). May, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Instructional planning: Information collected by school psychologists vs. information considered useful by teachers (Research Report No. 30). June, 1980.

Algozzine, B., Webber, J., Campbell, M., Moore, S., & Gilliam, J. Classroom decision making as a function of diagnostic labels and perceived competence (Research Report No. 31). June, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., Richey, L., & Thurlow, M. L. Psychoeducational assessment and decision making: A computer-simulated investigation (Research Report No. 32). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., & Richey, L. Psychoeducational assessment and decision making: Individual case studies (Research Report No. 33). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R., Potter, M., & Richey, L. Technical supplement for computer-simulated investigations of the psychoeducational assessment and decision-making process (Research Report No. 34). July, 1980.

Algozzine, B., Stevens, L., Costello, C., Beattie, J., & Schmid, R. Classroom perspectives of LD and other special education teachers (Research Report No. 35). July, 1980.

Algozzine, B., Siders, J., Siders, J., & Beattie, J. Using assessment information to plan reading instructional programs: Error analysis and word attack skills (Monograph No. 14). July, 1980.

Ysseldyke, J., Shinn, M., & Epps, S. A comparison of the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability (Research Report No. 36). July, 1980.

Algozzine, B., & Ysseldyke, J. E. An analysis of difference score reliabilities on three measures with a sample of low achieving youngsters (Research Report No. 37). August, 1980.

Shinn, M., Algozzine, B., Marston, D., & Ysseldyke, J. A theoretical analysis of the performance of learning disabled students on the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 38). August, 1980.

Richey, L. S., Ysseldyke, J., Potter, M., Regan, R. R., & Greener, J. Teachers' attitudes and expectations for siblings of learning disabled children (Research Report No. 39). August, 1980.

Ysseldyke, J. E., Algozzine, B., & Thurlow, M. L. (Eds.). A naturalistic investigation of special education team meetings (Research Report No. 40). August, 1980.

Meyers, B., Meyers, J., & Deno, S. Formative evaluation and teacher decision making: A follow-up investigation (Research Report No. 41). September, 1980.

Fuchs, D., Garwick, D. R., Featherstone, N., & Fuchs, L. S. On the determinants and prediction of handicapped children's differential test performance with familiar and unfamiliar examiners (Research Report No. 42). September, 1980.

Algozzine, B., & Stoller, L. Effects of labels and competence on teachers' attributions for a student (Research Report No. 43). September, 1980.

Ysseldyke, J. E., & Thurlow, M. L. (Eds.). The special education assessment and decision-making process: Seven case studies (Research Report No. 44). September, 1980.

Ysseldyke, J. E., Algozzine, B., Potter, M., & Regan, A. A descriptive study of students enrolled in a program for the severely learning disabled (Research Report No. 45). September, 1980.

Marston, D. Analysis of subtest scatter on the tests of cognitive ability from the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 46). October, 1980.

Algozzine, B., Ysseldyke, J. E., & Shinn, M. Identifying children with learning disabilities: When is a discrepancy severe? (Research Report No. 47). November, 1980.

Fuchs, L., Tindal, J., & Deno, S. Effects of varying item domain and sample duration on technical characteristics of daily measures in reading (Research Report No. 48). January, 1981.

Marston, D., Lowry, L., Deno, S., & Mirkin, P. An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study (Research Report No. 49). January, 1981.

Marston, D., & Deno, S. The reliability of simple, direct measures of written expression (Research Report No. 50). January, 1981.

Epps, S., McGue, M., & Ysseldyke, J. E. Inter-judge agreement in classifying students as learning disabled (Research Report No. 51). February, 1981.

Epps, S., Ysseldyke, J. E., & McGue, M. Differentiating LD and non-LD students: "I know one when I see one" (Research Report No. 52). March, 1981.

Evans, P. R., & Peham, M. A. S. Testing and measurement in occupational therapy: A review of current practice with special emphasis on the Southern California Sensory Integration Tests (Monograph No. 15). April, 1981.

Fuchs, L., Wesson, C., Tindal, G., & Mirkin, P. Teacher efficiency in continuous evaluation of IEP goals (Research Report No. 53). June, 1981.

Fuchs, D., Featherstone, N., Garwick, D. R., & Fuchs, L. S. The importance of situational factors and task demands to handicapped children's test performance (Research Report No. 54). June, 1981.

Tindal, G., & Deno, S. L. Daily measurement of reading: Effects of varying the size of the item pool (Research Report No. 55). July 1981.

Fuchs, L. S., & Deno, S. L. A comparison of teacher judgment, standardized tests, and curriculum-based approaches to reading placement (Research Report No. 56). August, 1981.

Fuchs, L., & Deno, S. The relationship between curriculum-based mastery measures and standardized achievement tests in reading (Research Report No. 57). August, 1981.